Figure 1

Anopheles gambiae odorant receptor 1 genomic sequence (SEQ ID NO: 9)

Features:
1) Presumed Untranslated 5' and 3' regions are underlined.
2) Potential TATA box transcription initiation signal is double underlined.
3) Putative Start (ATG) and Stop (TAA) codons are in BOLD.
4) Introns are tentatively assigned and are shown in lower case.
Exons are highlighted.

```
AGCTTTGTTCATTTATGTTGAAATCTAGCCCATTTTGTATAGTGCTGAACGACGAAGAACATACGAAAGTACCTCGT
CCGAACACTATCAACATTAATTATACCAAGCTAGAAGAAGATATTTATAGTCAAGCCTCAACATCATAGGAAACTTT
AGCAAAACCATTTAATTTACATGATGATAAGTCCCACCTCTTACCCCAGCACAGGTTTGAGAAGGACGAAAGTATCT
TTACGATAATATTACTCTAAGGTAGTTTTTGAATAAAATAAAAATTTACGTGCAAGTGGTGGCATCGGACATCATTC
GAAAGAATCTACTAAGTCATACACACACCCAAGACGACCGACGTAGTTTCATCTAGAAAAAACGGGTCAGCTCCATC
GAACACGTCAGGACATAACTGCGACATGCGTATGGTCAGTTCCACTAGTGCCAACACTGGTTCCAGGGCACTACCTT
CCGAAGCAGTAGAACCTAATGTATTGGAAATTATTAGGACATACTGCAACATGCATATGGCTAGTTCCGCTGGTACC
AACGATGGCACCAGGACACTATCTGCGGCCTTGTAAAATCACTGTAAAATCTATACAAAAACGGCTTTACCCATACT
TTATCACAAAAACGGCAGGTGAGGGCTGGATTGCTTCAAAGCATTAGAAATATATAATTTCAAAGTCCATAATCTCC
TTAAAAGATAGACAaCAGTAGAGAACACATTTAGTGCTCTTTTCGTTCGAGTTAGTTGCCTTCTCAAGTAAGCGTTT
AATGCTCAATTGTTGTAGATTCGTTGGATGACTCTCGCTACGTGCTATAGTGGTCAATACTTCCAATTAGATTTCAT
AATTAGTTTCCAATTGTCCACGGAAAACCCaCAAAAGAAAAAAAAACTTGTATCTAGGGTGGAATTTTTCGAGAACA
ATTGGACACTTCATATGAAAAAGGACAGCTTTTTCAAAATGTTAAATAAACACCGTTGGATCCTTTgttggatttca
attctccaaattctgcagaataattctgcaaattttacaaaactgctcaaccaccaataattccaattaatcatctg
aacatttaaaactgataattaagatgagtaattgcttcgtcatcacctaagaaatcgattagtttggataaaaagaa
caaattgaaatacaataaagtccctgaattttattcgaataacggcttgaactcatttatttcaaaaacctttgaga
aattcctcgttgaaaattggtctcctatagttctgctaacgggccacttcaaaagcaagaactaacaaaatcataat
tatggtgcaagtaactatcagtaccagtaatcgccattaaaaacttttcctcaatttgcggctcgttaccggctaaa
tacagagcagagtaacgggaagtgatcaacgtcgctattagtataacgaggaacgccctccgaaggtgtgttgaagg
accttttcaaattgaaaccaagtactgtttccagtttttaaattggatagttataaaatgagccgttcaacgatcggg
catcatttgagtttcatcttcgaggagaaatagatcagtgccactgtttaaccgaaagtaatgaagctgaacaaact
gaacccacggtgggatgcgtacgatcgacgggattcgttctggttgcagttgctttgtttgaaatatttagGCCTAT
GGCCACCGGAAGATACGGATCAGGCAACGCGGAACCGGTACATCGCGTACGGTTGGGCTTTGCGGATCATGTTTCTA
CATCTGTACGCTCTAACGCAAGCCCTATACTTCAAGgATGTGAAGGATATTAATgtgagtctctagttagctattag
tgttccacctgtccataatctgtctttttattgggtagGACATCGCAAATGCATTGTTCGTGCTTATGACTCAAGTGA
CGTTGATCTACAAGCTGGAAAAGTTTAACTACAACATCGCACGGATTCAGGCTTGTCTGCGCAAGCTTAACTGCACA
CTGTATCACCCGAAACAGCGCGAAGAATTCAGgtaagcctgctgggaaatatgactaaaaagagtgctaacaaacga
ctctcctccaaatgtagCCCCGTTTTACAATCGATGAGTGGAGTGTTTTGGCTGATGATCTTTCTCATGTTTGTGGC
TATCTTCACCATCATCATGTGGGTTATGTCGCCAGCCTTCGACAATGAACGTCGTCTGCCCgTGCCGGCCTGGTTCC
CGGTGGACTATCACCATTCGGACATAGTGTACGGTGTACTGTTCCTGTATCAAACCATTGGAATCGTCATGAGCGCA
ACGTACAACTTCTCGACCGATACCATGTTTTCCGGCTTGATGCTACACATAAATGGACAAATTGTGCGGCTTGGTAG
TATGGTTAAAAAGgtgagttacggcgactacttgcctccagtaaggacagggagtttgtttccgttatgatatcatt
ttatcagCTTGGACATGACGTCCCTCCCGAACGCCAATTGGTCGCAACGGATGCGGAATGGAAAGAGATGCGAAAGC
GCATCGACCATCACTCCAAAGTGTACGGTACGATGTACGCTAAAGTAACGGAGTGTGTGCTGTTTCACAAGGACATC
TTAAGgtacgaattgggccaattaattgtgtcatttaaaaagcttgacccaacttttcacagcttcggcgatgaagt
gcaggacattttccaagGATCTATCTTCGCGCAAGTATGCGCGTCTGTAATTATCATTTGTATGACACTGCTGCAAC
TACCGGGGCGATGTTACGATGGCCGATCTGCTGGGCTGTGGGGTCTATTTGCTAGTAAaGACATCGCAAGTGTTTA
TTTTCTGTTACGTAGGGAATGAAATCTCCTATACGgtaggttggacacgtagaggaattaaatgtttgggaagaata
tcaataccaaatagtatgatgtttcgttacagACGGATAAATTTACAGAGTTTGTTGGGTTTTTCAACTACTTCAAG
TTCGATAAGCGTACCAGCCAAGCAATGATATTTTTTCTGCAAATgtgagatagcggtgtatttgtgcagtcagtaca
```

ttaaatacgttctctatttcagCACTCTTAAAGATGTTCACATCAAGGTGGGAAGTGTCTTGAAGGTTACGGTAAAT
CTTCACACATTTTTGGACgtatgtaattatgctgtggtatttagcttgaaataagctacaaactttgaaagtaattt
caatctgttttgtagATTATGAAGCTATCGTACTCCTATCTGGCCGTACTTCAGAGCATCGAATCAGAGTAATGGtC
CTAATATCCCTAATGTTGAAATTATATTTTGTTAGATTTATTGCATAAAGTAaTaTTTAATTTTATACATCAAACGT

5   AAGCCCGCtaGTTTTCAATTAGCCTTTTCCAAAATTTATCAAATTGATTTCGAATTGATTGCAGAGTTTCAGGAATT
TAATCTGATAGGATATCTTGTTTATCCAATAGAGGTGTGGAAGCGTTCCCAAGCCATTCGTTTGATAGTTTATAGCA
CCGTCGAGCAGTTGATCGCTGTGATCGCTAGGCGCACCTGATTTTATCTTTATCTCGCACCTGTTATGGCAAGGGCG
CTTTTCACACGTTTCACACAATATAATGCACATGTATAATGCATTCTTACTTTAGCATTTTTGTTACATATAATACC
AAAATTATGCATTTTTATTCTCACGCAACGATTAGAGGATGACTTcACAAAGGTCCATCTAGTGGTAGGAGGTATAC

10   AATTATACCTCTCAAAATCTCACAGCAtAATGAGAAACAAAAGGATACCAAGCATACCCTTTTTTTACTTGACAATT
TCATTTGATTTATGTAATAAAGCACTGCaCGTCGACTTCCTAAAA

Figure 1 continued

15

20

Figure 2

Anopheles gambiae odorant receptor 2 genomic sequence (SEQ ID NO: 10)

Features:
1) Presumed Untranslated 5' and 3' regions are underlined.
2) Potential TATA box transcription initiation signal is double underlined.
3) Putative Start (ATG) and Stop (TAA) codons are in BOLD.
4) Introns are tentatively assigned and are shown in lower case.
5) Exons are highlighted.

```
GGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTCCCTCACCGTGACGTGCTAGAAATGGTTCAACATACTCGT
CCGGCAGAGCGAAGACGACGAACAGCGGAATGTCCCAGGAAATGTAATGAGATATCACAGCAAGTGAACCCAAACCG
AGCTGTGCGCTTTGTGTTGCGCTTTAAAAATGGCCCTTCCTTCGCCGCATCTGCTTGGTTTCACACGCTTTCCCAGG
AAATCCACTGACCACTGGCCACACATCAACCACCGGAGCGGGAGCCTCAGTGCCCAGCGAAGCATATAATTTGCTCA
AAAAGTCACGGTACTCAATTAATTTGATTATAATCAATTTCGTGGCTTCCAACACACCCTTCTTCCACAATCCATCG
CCGAGTGAGCGAGTATAAAGGTGAAGAAACGTACCTTGCGCTTGCTCACTAACTGAACCGGATTTCAAAAAGGAACA
TAAACCGCAACCCACAGCCGAAAATGCTGATCGAAGAGTGTCCGATAATTGGTGTCAATGTGCGAGTGTGGCTGTTC
TGGTCCTATCTGCGGCGGCCGCGGTTGTCCCGCTTTCTGGTCGGCTGCATCCCGGTCGCCGTGCTGAACGTTTTCCA
GTTCCTGAAGCTGTACTCGTCCTGGGGCCACATGAGCCGAGCTCATCATCAACCGATACTTTACCGTGCTGTACTTTA
ACGTCGTCGtacgtgggcgaggggaggggcaataaccttcccacttggtggatattttcataccttttccatgtgtt
tttttattctctgtttgttgccatccagCTCCGAACCTGCCTTTCTCGTGATCAATCGACGGAAATTTGAGACATTTT
TTGAACGCGTTGCCCGCCGAGTACGCTCTCCTCGACgtaagtcattggttttttctagttttttgggggagttgtttaca
ccataaccacccccgacggtaacatttgatcgtcccgcgaaaatgtttgtacagAAAAATGACGACATCCGACCCGT
GCTCGCAGCGGTACACACGCGCGGGGACGGCATGCTTATCGATATCGAATCTGTGGCTCGGCGCCTTCATTAGTCCGTCGC
TTCTGACCTATCCTCTCTGTTTGTGCCGCGGGCCGCGGCCTACGGTACGGCGTCACGATACCCGGCGTGGACGTGCTGGCC
ACCCCGACCTACCAGGTCGTCGTTTGTCCTGCAGGTTTACCTTACCTTCCCGGCCGTGCTGCATGTACATCCCGTTCAC
CAGCTTCTACGCCGACCTGCACGGCTGTTGCGGCTCGTCCAGATAGCGGCCCTAAAGCAACGGCTCGGCAGCGCTTGGGGCC
GCCACAGCGGCACGATCGGCTTCGACCGGACACAGCGCCCGCACACTGTTCGCCGAGCTCAAGGAGTGTCTAAAGTAT
CACAAACAAATCATCCACtaagtagacgctagtagactcgaccggattgcccttccctcggggaggggaggtttgct
atttcgggatgcggcagcacgcatacacacaaaccggaagccattaattctcccgtttttcatgcccgcacgggcact
gggtcatgtttcacatccttccttcctttccaaacacacacacgcgcgcgtgcacgtacagATATGTTCATGATCTC
AACTCACTCGTCACCCATCTGTGTCTCTGCTGCAGTTCCTGTCCTTCCGGATCATGCTGTGCGCACTCCTGTTTCTGCT
AAGCATTGtaagtaaaatcgaccgacgtgcggtcgctagtccgtctccggactctcatttcgggactcaatcgttcc
atctctcaatagAGCAATCAGCTGGCACAGATGATAATGATTGGATCGTACATCTTCATGATACTCTCGCAGATGTT
TGCCTTCTATTGCCATCGCAACGACGTACTGGAGCACGtaatggcgctgaagctgagtttggttgagcggttcgcta
tagatcggctgtcttacattgttgtgtttctgcatggggatcggttttgttttttcctctccatttcagAGCCTAGCG
ATTGGCCATGCCATTTACAATGGACCGCTGCCCGGACTTTGAGGAACCGATAAGGAAACGGTTCATTCTAATTATTGC
ACGTGCTCAGCCGACCGATCGTGGTAAGtttggctgatcgatgctctgttcaatgaacatggcacagaaggctgtgta
aatagctgttcattaataagtttttttcagaatgtatcgttttttagttgatttaaacgcattgttctatgcaatggta
gcaacaatagaccgccctttattaatccaagcttcctttaggattgatttttattttaagagaaagataaaccatttt
tagtaaccaatttagttacaggaaccaaaatacagaatttattattattattattattattattattattattatta
ttattattattattattattattattattattattattattataattattattattattattattattatt
attattattattaatattattattattattattattattactattattattataattattactttttattattat
tattattattattattattattattattattattattattattataattatgattattattattatta
ttattattattattgttattcattattatacattattatcataataataattttattatgattattattattattatta
ttattattattattattattattattcttattattattattattattattattaatattatttttaatatt
attattattattactattcttattataattattttttttattattattattattattattattattattatta
ttattattattattgctattgttattattattcttattattgctattgttattattattattcttattattgttgtt
gttgttgttcttattattgttgttgttgttattcttattattgtttattattattgtttttttttttattctctaatta
```

3

```
       ttccagtaatccataataaaaaataataaagtaaataaatagtaaatagtaaataattccagtaactgtagtaatac
       acaataatctctaagaattaaaattgcattttgtaatgaaatatgttgattgttcgaatagttcagaaaaacttaaa
       aatgcctcagcattaaacagttttgaggttgttcagggcatttagtttagatattttagtattttaaagcatttgtt
       ttcattactacaaaaaagcaaatttatgagtgaattactttcagttcttctaaacgcctatgtgtatgcaattacat
   5   aacaatagctctctttttttattgcattttttccttagtaatctaaatccaatctcttctttccctcttgcagATTAAA
       GTCGGCAACGTGTACCCGATGACGTTGGAAATGTTTCAAAAATTGGTCAACGTGTCCTAGTCGTATTTCACACTGCT
       GCGCCGAGTGTACAACTAAACTTAACCGGTAAACAAACAAAAATCCCCTCATCACTATGCAAAGACAGCAAGCAGCC
       GATCATCAAACACCATTAGCAGCCACAAAGTTACCAGCCGCTTATCCCACGGGATTTGGTGGAAAGTTATTGCACTG
       AAGCTCTTTCACCCAAATTTTCATGGAGGTTCCCTCTCAACCAACCCATTGAAGCGAATAAAAGTATCAGCAACCAG
  10   GCGACGGTGAAAAAAACGCTGCATTATTGTGCTTGCTTCAGCATTCCAGCGAATGACTCTTAAACTTTTCCATTCAAA
       AGTCGCGATGCTCACGATACGGAGCGGTGTGTTGTTCGATCCGCCGAGTGCACTCGCAAGCCGGTGATGTTGCCGGT
       GGAAATGCACAGATCGACACAGCGATAGATAATCGTTTGTTCGCGTAAATGGGAGGGAAAAAAGTAAGCTGCCAGCT
       ACTTCATTTCCATGTTAATTGAAACTCAAGCCAACGAACATGCAGAACCCGGTTGGTTGTGTGTCTCCGCTCCGGGA
       AAGGTCTCTGCTCCGGGGCATGGATTCTTTCCCCCTCCGGGTGGTTGGGGGTATTGTTTAGGTTTTTATTTTACAAA
  15   TTCATATCCTTCCGCTTCCGCATCAGCCGACCCGGTGGGTGCGCCAGACAGATGTGCGGCGGGCAACAAAACTATGC
       ACGAACATGGCCAACAAACACAGCTTCTATCTCATCTCTGTGTCGCACTGTCTCGCTTTCCCGCTGCGTTGCTTGTA
       GTACTATCATTGTTTTAGTCCACGGGTTTACTTCTAATTCCATTGCACCACGCAAAAAGGCTCATCCTTTGCTCGTT
       CCGGTTGCAACTTCGACAAGCGCATGGTTGGGATACGAACAAAAAACCAACTACTCCACCCACTACTACTACTACTG
       CCACCACCACTAACAACACTACACTTGGTTGGGAGCTTGCAGACCCACAAGCAAACAACGATACAAGCTAGCTAGCT
  20   GCTGTGTGCGCTCGAGTCAGCCGACGGTACAAGGTTTAACCGGTACAAGCAACTCCCGGACCGATCCCAAAACTCTG
       ACAAGGCACGGGGCCGCATCCGGCAGTACGGTCGGAAAACATGGAAATGTTTAATTAAAACTGTAATTGTCAATCGC
       TGCTACAAGTTGTGACACAGGGAGAGAGAGAGACAGAGCGCGCCCGATGGTGATGGTGTAAAAGATAGATACAGGAA
       AAGAGCGAGAAACATTGGTACGATTTGGTGTGGTTAGCAAATTTGATTTCCACTGATTTTGAGTGCAAATTTAATGC
       ATCGAAAATTTGCCATTCAGGGTAAAGTTGCTCGTGGACGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTAT
  25   CGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGGA
```

Figure 2 continued

30

Figure 3

Anopheles gambiae odorant receptor 3 genomic sequence (SEQ ID NO: 11)

5

Features:
1) Presumed Untranslated 5' and 3' regions are underlined.
2) Putative Start (ATG) and Stop (TAA) codons are in BOLD.
3) Introns are tentatively assigned and are shown in lower case.

10   4) Exons are highlighted.

```
     AAGCAGAACACATCAAGAAGCAATTAGGTGTGTCGTACGTTAGCAAGTAGTTCGCGAGGAGGAATAAAATAGATGCC
     TTCTGAGCGGCTTCGTCTCATTACTTCCTTCGGAACTCCTCAAGACAAACGCACGATGGTACTGCCAAAATTAAAGG
     ATGAAACAGCAGTGATGCCGTTTCTGCTGCAAATTCAAACCATTGCCGGACTGTGGGGTGACCGTTCCCAGCGGTAC
15   CGTTTTTATCTCATCTTTTCCTACTTCTGCGCGATGGTGGTTCTACCCAAAGTGCTGTTCGGTTATCCAGATCTCGA
     GGTTGCGGTACGCGGCACGGCCGAGCTGATGTTCGAATCGAACGCATTCTTCGGCATGCTAATGTTTTCCTTTCAAC
     GCGACAACTACGAGCGATTGGTGCATCAGCTGCAGGATCTGGCAGCTCTAGgtgagtatgcagccaatcgattgttc
     caaaccttcgcaacatccttcgtaacactgctacactttcagTCCTCCAAGACCTACCCACAGAGCTGGGAGAGTAC
     CTGATCTCAGTGAACCGACGGGTCGATCGGTTCTCCAAAATTTACTGCTGCTGTCACTTTTCCATGGCAACGTTCTT
20   TTGGTTCATGCCCGTCTGGACGACCTATTCCGCCTACTTTGCTGTGCGCAACAGCACGGAACCGGTCGAGCACGTGT
     TGCACCTCGAGGAAGAGCTGTACTTCCTGAACATTCGGACTTCGATGGCGCACTATACGTTTTATGTGGCCATTATG
     TGGCCCACGATCTATACGCTCGGGTTTACCGGTGGCACAAAGCTGCTGACCATTTTCAGCAATGTTAAGTACTGTTC
     GGCCATGCTGAAGCTCGTTGCACTCCGAATCCACTGTCTAGCGAGAGTAGCGCAAGACCGAGCGGAAAAGGAGCTGA
     ACGAGATTATTTCCATGCATCAGCGGGTACTCAAgtaagtaaattcaaattgaaagttttgcagggaataacttgag
25   tgtgtctgacccgtgcacatcctagCTGCGTGTTCCTGCTGGAGACGACATTCCGCTGGGTATTTTTCGTGCAGTTC
     ATTCAGTGTACAATGATCTGGTGCAGTCTCATCCTCTACATAGCGGTGACGgtaatagcattttcgtcatttcgtta
     gccttattcaatccatttttgtgaacgtgaatttcccccagGGGTTCAGCTCGACGGTAGCGAATGTATGTGTCCAG
     ATCATTTTGGTGACGGTGGAAACTTACGGCTACGGCTACTTCGGAACAGATCTAACCACGGAGGTGCTTTGGgtacc
     ctttggatgaagcttcaaaaagtaattccaaattctgtttttcgatttttccccttttccactagAGCTATGGCGTTG
30   CCCTCGCCATTTACGATAGCGAGTGGTACAAGTTTTTCCATTTCGATGCGCCGCAAACTTCGACTGCTACTGCAACGA
     TCCCAAAAACCGCTCGGCGTAACGGCGGGAAAGTTTCGCTTCGTCAATGTGGCCCAGTTTGGCAAGgtaacattaat
     tacagtttgaaaattctgaagaatgcatcttacttgccttacttgttgttccagATGCTCAAGATGTCCTATTCATT
     TTACGTAGTACTGAAGGAGCAGTTTTAGGAGCTGCTGTTTCCCACCCTGGAAATGGCCTTTTCGCACTGTCTTCTGT
     TTGTTGGACGCACGCAGCACCGAGAGCGCCCCTGCACGCACTGACGTATTTTGGCTACTTTGACGTTTGCACCTTTG
35   ACAGCTGAAGGACAGGGTACAATTTTTGCTGCTGTTATTACGCGCAGCGCATTGGATACGAAAACATTGGCCACAAG
     TTCTACGATTTTAGCGTTTATTTACTGTTCGTAGCAGCTTTTTTCCaCAATAAACACACACAATAACGTACCGACAG
     TATTCTTTTCATTGTAGGATAGAGAAGCCGCCGGCCAGCAGCCAAAACGCGCCGCAAAACGAAAGGCGGCACCACCG
     GGGGAAAAACACGGGAGCAAAACGAGAACAGAACGCAGTAAACAACAAAACCGGCCGGAACAACAACGGTGCCGGAA
     ACGA
```

40

Figure 4

Anopheles gambiae odorant receptor 4 genomic sequence (SEQ ID NO: 12)

5

Features:
1) Putative Start (ATG) and Stop (TAA) codons are in BOLD.
2) Introns are tentatively assigned and are shown in lower case.

10   GGGGAACTCCCCCACCCGACCAGACGACGGAAAGCTAACGATGTGCAATTGAATAGTCATTAGT
     AGCGTTTTTGCTCGCAAACGAACTAACCCTTTGACTTTTTAAGTTCACTACGGTGAGGACAAAAA
     TCAATAAATTAAATCGAGACCGTTGATGAGCAAAAGAAAAAAAAATATTTTACTGATTTTCATTT
     CGTTCCATCGACTACATAATCATAATTATATGCCACATTTTATTATAAGTTTTTGTATCATTTTTA
     AACAACACAAAAATGCATCCTTTCGAATATTAGTCAGGTTGTATCAACAATGAAGTTTGAACTGT
15   TTCAAAAATATTCCTCCCCGGACACGGTCTTATCCTTCGTGCTAAGGCTTTTGCATATCGTGGGC
     ATGAATGGGGCAGGATTTCGGTCGCGAATTCGAGTTGGTGGCATTTTTCTGTTCTATTTAATCTT
     TCTTGTAATACCGCCACTAACGGGCGGGTACACCGATGGTCACCAGCGTGTACGCACCAGTGTG
     GAATTCCTGTTTAATTGCAATATTTACGGCGGCAGTATGTTCTTTGCCTACGATGTGGCCACTTT
     CCAAGCGTTCATCCAGGAACTGAAGAGCCTTTCGGTTTTGGgtaatatttaattaattaaaattgcgtttattgcat
20   catcatttgtttctctttgcagTATGCTCACATTCGTACAGACTAAAGTATAAGCTGACCCGGTTCAACCGTC
     GAGCGGATATTATCGCCAAAGTGCAAACGACCTGCATGGGTGCTGTAACGCTTTTCTACTGGAT
     TGCACCGATACCTTCCATCTGTGCGCACTACTACAGGTCGACCAATTCCACCGAACCCGTGCGG
     TTTGTGCAACATTTAGAGGTGAAGTTCTATTGGCTCGAGAATCGCACCTCAGTCGAGGACTACAT
     AACCTTCGTGCTGATCATGCTACCCGTCGTGGTTATGTGTGGTTACGTATGCAATTTGAAGGTGA
25   TGACCATCTGCTGCAGCATTGGACACTGTACACTGTACACCAGGATGACTATAGAGATGGTAGA
     GCAGTTGGAAAGCATGGCATCAGCGGAACGAACTGCCAGCGCCATACGCAACGTGGGGCAGAT
     GCACAGTGGTTTACTGAAATGCATTAGGCTTTTGAACACGTCAATCCGATCGATGCTGATGCTGC
     AGTGGTTGACCTGCGTGTTAAACTGGAGCATTTCTCTCATCTATCTAACGAACGTGgttagttttgtctt
     gtttggaaatccaaaaacaaaaagatggctataattgaactttctattacagGGCATCTCGCTACAATCGGTTACCGTGGT
30   GGTAATGTTTTTTCTTGCCACTGCGGAAACTTTCCTGTATTGTTTACTTGGGACGCGGCTTGCGA
     CACAACAGCAGCTGCTGGAGCACGCACTCTATGCTACACGGTGGTACAACTACCCAATAGCCTT
     TCGCAGCAGCATTAGGATGATGTTGAGACAGTCGCAAAGGCATGCACACATAACGGTGGGGAAG
     TTTTTTCGCGTTAATTTGGAAGAATTTAGCAGGATTGTCAACTTATCCTACTCTGCTTACGTCGT
     ACTTAAGGATGTAATAAAGATGGATGTACAGTGAATGTTTTTTTTTTGGCTTGGCAACGAATGA
35   AGTTTTCCGAATCTATATTAGATCTAGAATTTAATCTAGATGTCATAATATGATCTTGGCCATGA
     CCGGTTCCTGGTTTTGGAACCAATTCTCAAAACAATTTTGAACTTAGGGCGAGGCATGAAATGTC
     CCAAGAACCTATCCAAGTTCTGGAACTACATATTACCGAATCTATCCCATTATTGCCTCGGAACT
     GGTTTGGTGCTAAATATTTGTCCAAATGTTGGTCCTGGACCTATCCAGACAAAGATCTTCAATTA
     TTCCTACCACTGGAACTGATTAATTGATGTAGGAAGTCATGGAGGTGTTCAGGGAGAATTTAAA
40   CACTAATGTTCCAACTCATTATTTCAAGGGCAATTCTATTTTTTATATGCCCCTACGGATTGATAC
     GTATGTATTACTCCATTTCCTGGACTTTGTCTTATTCTTGCTGCTGATTGGACGTGAAATGTTGA
     GAAAAAGATTCTTATTTATGAGTGATACAGAGCCTTTAAATACTCCTACGTTGTTTGCTATTTAA
     GTATGGCCAGGCTAATCACAATCGCTACTAATGAACAGAATCTCTTCTAATTAAACCCTTTCGAT
     TGATAGTGTCAATGTCAATGTCGAGATAATTGAACTGCAAACgATACCTACCTTAAACGGAGCAG
45   AACACATCAAGAAGCAATTAGGTGTGTCGTACGTTAGCAAGTAGTTCGCGAGGAGGAATAAAAT
     AG

50

Figure 5

ANOPHELES GAMBIAE

Preferred DNA Codons

Amino Acids			Preferred Codons					
Alanine	Ala	A	GCC	GCG	GCT	GCA		
Cysteine	Cys	C	TGC	TGT				
Aspartic acid	Asp	D	GAC	GAT				
Glutamic acid	Glu	E	GAG	GAA				
Phenylalanine	Phe	F	TTC	TTT				
Glycine	Gly	G	GGC	GGT	GGA	GGG		
Histidine	His	H	CAC	CAT				
Isoleucine	Ile	I	ATC	ATT	ATA			
Lysine	Lys	K	AAG	AAA				
Leucine	Leu	L	CTG	CTC	TTG	CTT	CTA	TTA
Methionine	Met	M	ATG					
Asparagine	Asn	N	AAC	AAT				
Proline	Pro	P	CCG	CCC	CCA	CCT		
Glutamine	Gln	Q	CAG	CAA				
Arginine	Arg	R	CGC	CGG	CGT	CGA	AGA	AGG
Serine	Ser	S	TCG	AGC	TCC	AGT	TCT	TCA
Threonine	Thr	T	ACG	ACC	ACT	ACA		
Valine	Val	V	GTG	GTC	GTT	GTA		
Tryptophan	Trp	W	TGG					
Tyrosine	Tyr	Y	TAC	TAT				

5

http://www.kazusa.or.jp/codon/cgi-
bin/showcodon.cgi?species=Anopheles+gambiae+[gbinv]

Figure 6

Name	SEQ ID NO
Arrestin 1 (cDNA)	SEQ ID NO: 1
Arrestin 1 (polypeptide)	SEQ ID NO: 2
Odorant Receptor 1 (cDNA)	SEQ ID NO: 3
Odorant Receptor 1 (polypeptide)	SEQ ID NO: 4
Odorant Receptor 2 (cDNA)	SEQ ID NO: 5
Odorant Receptor 2 (polypeptide)	SEQ ID NO: 6
Odorant Receptor 3 (cDNA)	SEQ ID NO: 7
Odorant Receptor 3 (polypeptide)	SEQ ID NO: 8
Odorant Receptor 4 (cDNA)	SEQ ID NO: 13
Odorant Receptor 4 (polypeptide)	SEQ ID NO: 14
Odorant Receptor 5 (cDNA)	SEQ ID NO: 15
Odorant Receptor 5 (polypeptide)	SEQ ID NO: 16
Odorant Receptor 6 (cDNA)	SEQ ID NO: 17
Odorant Receptor 6 (polypeptide)	SEQ ID NO: 18
Odorant Receptor 7 (cDNA)	SEQ ID NO: 19
Odorant Receptor 7 (polypeptide)	SEQ ID NO: 20

Figure 7

Anopheles gambiae odorant receptor 5 genomic sequence (SEQ ID NO: 21)

Predicted Exons: *ITALICIZED*, <u>UNDERLINED</u> AND HIGHLIGHTED.
Introns: lowercase.

tctagacttgaacccatgacgggcattttattgagtcgttcgagttgacgactgtaccacgggaccacccgtttatcactatcactatt
aattaattataatatgcttttgtagcgatcagcctaccgggttttgtttctctggatatcttaagttcccatttgattatcaagatagaa
caacaacttgtaccttaaataatcattacgtacccttaatcaacctgtgcatcaaggagttttcgcgaaagcaaaaatccgattgtct
gatgttgtcttgattccatccgattcgttactggttctgcaaaatcgtccaataatacggcaatgtccttatcgatgcttgaatcaacat
cacattgtttgcatttcgtttttttgcgtgcaaatatgttatttgcaaagaaggcaaggtaatgtgcttaagagtaaatacaattcgctg
tccattttttgtccaccagtgtgccagaacccgtgccttttagtccttcgaatacatccgaccagtcagcaagcaagtgcatc*ATGG*
TGCTACCGAAGCTGTCCGAACCGTACGCCGTGATGCCGCTTCTACTACGCCTGCAGCG
TTTCGTTGGGCTGTGGGGTGAACGACGCTATCGCTACAAGTTCCGGTTGGCATTTTA
AGCTTCTGTCTGCTAGTAGTTATTCCGAAGGTTGCCTTCGGCTATCCAGATTTAGAGAC
AATGGTTCGCGGAACAGCTGAGCTGATTTTCGAATGGAACGTACTGTTTGGGATGTTG
CTGTTTTCTCTCAAGCTAGACGACTATGATGATCTGGTGTACCGGTACAAGGACATATC
*AAAGATTG*gtgcgtgataatgattgataaaaggaacctttgagcaactcctatccctttcaag*CTTTCCGTAAGGAC*
GTTCCCTCGCAGATGGGCGACTATCTGGTACGCATCAATCATCGTATCGATCGGTTTTC
CAAGATCTACTGCTGCAGCCATCTGTGTTTGGCCATCTTCTACTGGGTGGCTCCTTCGT
CCAGCACCTACCTAGCGTACCTGGGGGCACGAAACAGATCCGTCCCGGTCGAACATGT
GCTACACCTGGAGGAGGAGCTGTACTGGTTTCACACCCGCGTCTCGCTGGTAGATTAC
TCCATATTCACCGCCATCATGCTGCCTACAATCTTTATGCTAGCGTACTTCGGTGGACT
AAAGCTGCTAACCATCTTCAGCAACGTGAAGTACTGTTCGGCAATGCTCAGGCTTGTG
GCGATGAGAATCCAGTTCATGGACCGGCTGGACGAGCGCGAAGCGGAAAAGGAACTGA
*TCGAAATCATCGTCATGCATCAGAAGGCGCTAAA*gtaaggtctgccggtatgttgtggatagaatacattt
ctagctgctttcag*ATGTGTGGAGCTGTTGGAAATCATCTTTCGGTGGGTTTTTCTGGGACAG*
*TTCATACAGTGCGTAATGATCTGGTGCAGCTTGGTTCTGTACGTCGCCGTTACG*gtaacta
aaagcactgtagtgatctgtctgccacaccattcactgctgtgtcttgttttgtcactcttcccag*GGTCTCAGCACAAAAG*
CGGCAAACGTGGGTGTACTGTTTATACTGCTAACAGTGGAAACCTACGGATTCTGCTA
*CTTTGGCAGTGATCTTACCTCGGAGGCAAGTTGTTATTCGCTGA*gtttcagttacttttccgttcccc
tctaaccgtaccacttgtaccatttgtttgagacagagcttgagcgtag*CACGTGCTGCGTACGGTAGCCTCTGG*
TATCGCCGTTCGGTTTCGATTCAACGGAAGCTTCGAATGGTACTGCAGCGTGCCCAGA
AACCGGTCGGCATCTCGGCTGGGAAGTTTTGCTTCGTCGACATTGAGCAGTTTGGCAA
*TG*tatggggagaccttccactgtggcaagaaagattttctttattaatgcatcttttaatttacag*ATGGCAAAAACATCA*
*TACTCGTTCTACATCGTTCTGAAGGATCAATTTTAA*aggggaactcccccacccgaccagacgacggaa
agctaacgatgtgcaattgaatagtcattagtagcgttttgctcgcaaacgaactaaccctttgactttttaagttcactacggtgag
gacaaaaatcaataaattaaatcgagaccgttgatgagcaaaagaaaaaaaaatattttactgattttcatttcgttccatcgacta
cataatcataattatatgccacattttattataagtttttg

Figure 8

Anopheles gambiae odorant receptor 6 partial genomic sequence (SEQ ID NO: 22)

5   These are the predicted last three exons of another candidate *Anopheles gambiae* odorant receptor.

Predicted Exons: *ITALICIZED*, <u>UNDERLINED</u> AND HIGHLIGHTED.
Introns: lowercase.

10

aacacccatcttatcggcaaaattagtatttaccgtttgaaagcggcttcccttcctggctgtttctcactctctctctctctgtctctctta
ttgatgccgtatgcgccgcgtgctataggctag*TTATGCTTACCGGATGTTGCGATCGCGCACGTGCTTT*
TCCGCATACGCCAGTGCACACTTGATGGCGGTGGTGATGACGTCTGCTGCGCACCGTT
15   *TTCTGCTCGTGAGTCAGACCTTTTCATTTCCTGCAATATCCTGTTTCTTTCCCGACCCC*
ACAGACGGTTAGACGGATATATGCTGGTAAAGTTTGTCCTCTTCATGCTGTGCTTTCTG
*ATCGAGCTGCTGATGCTGTGTGCGTACGGTGAGGATATTGTGGAATCG*gtaaggcaccaggc
ggtgatgagcgagtcgcgagtaattgaagcttttgcttttaaaacacatcagag*CCTTGGGGTGATTGATGCCGCT*
TACGGTTGCGAATGGTACCGGGAAGGGTCGGTGGCGTTCCATCGATCCGTGCTGCAAA
20   *TTATACACCGCAGCCAGCAGTCCGTCATACTGACCGCATGGAAAATTTGGCCCATCCAA*
*ATGAGTACTTTCAGTCAG*gtgagttgccaattgattgccgtttgcgttaatatttcagtaagagtgcgctctttcccttag
*ATCCTGCAAGCTTCCTGGTCCTACTTTACCCTCCTGAAGACCGTCTACGGGAATAA*gtaa
gcgcgagagagagagagagagcagtatcgttcaccctttggatgaatcaatagatttctaatcatgaaccattgaaaaatgaatca
acattttcgctagttgcacaatattgtaccattctatacagcttcaccacgaccaagcgtttgttgcatcaggaccaaacacgtttcga
25   caagccgcgtcacctgctggc

Figure 9

Anopheles gambiae odorant receptor 7 genomic sequence (SEQ ID NO: 23)

5

Features
1. Predicted Exons (7): *ALL CAPS, ITALICIZED, UNDERLINED, HIGHLIGHTED*
2. Introns (6): lowercase
10 3. 5' and 3' sequences: lowercase, dotted underlined

ccgcccgggcaggtgacttacgcggtctgacttgctggtgcgctgctttgtacggcaaacggctacacaagcgaatcgaattattttcc
tatcacgctgcgcttaccagcgcctgctggtaggcaaagaatgtgcaaagtttcatttggcttggttcgtctgctttgctgtgaacgtgt
15 gcacggttgcatcgctaaggtttcggtgtgagccgagaagttgcagatcgaaatctctttgtgtgtgtgtgtgtgtgtgcagtgggaa
gcattgtgtttagtgagaagtgaaaagaaaagtgctgaaaaatgcaagtccagccgaccaagtacgtcggccttcgtttgccgacct
gatgccgaacattcgggttgatgcaggccagcggtcaactttctgttccggctacgtcaccggcccgatactgatccgcaaggtgtac
tcctggtggacgctcgccc*ATGGTGCTGATCCAGTTCTTCGCCATCCTCGGCAACCTGGCGACGA*
ACGCGGACGACGTGAACGAGCTGACCGCCAACACGATCACGACCCTGTTCTTCACGGCA
20 *CTCGGTCACCAAGTTCATCTACTTTGCGGTCAACTCGGAGAACTTCTACCGGACGCTC*
GCCATCTGGAACCAGACCAACACGCACCCGCTGTTTGCCGAATCGGACGCCCGGTACC
ATTCGATTGCGCTCGCCAAGATGCGGAAGCTGCTGGTGCTGGTGATGGCCACCACCGT
*CCTGTCGGTTGTCG*gtatgtgtgtatgtgtgtggccgtttgggaaagtgtctttgcggcagaaccccaatctactgttacgc
ttgactgggttttgtttttttctcggtggagggacgggataaaatatctgaaagaataattgagtcaacccacaggggggatgcaag
25 acatcgcaggcagagagtttgggtttgatttatcaccgcacaccgaatatcttcacggttcataagcttcaccgcggtgaaaaggga
actcccatttccctgttttcttttttttcttcctctcgataaaattactcatcgcttttcgtttttttttttttgttgttgcttctttcttctttcatc
cctactag*CCTGGGTTACGATAACATTTTTCGGCGAGAGCGTCAAGACTGTGCTCGATAAG*
GCAACCAACGAGACGTACACGGTGGATATACCCCGGCTGCCCATCAAGTCCTGGTATC
CGTGGAATGCAATGAGCGGACCGGCGTACATTTTCTCTTTCATCTACCAGGTACGTTG
30 *GCGGAAT*gtcctgcgcgtcacagttggcagtcagtgagcggcaacacggcgaaaaaatgggactaaaaccggtcttcacaga
gccaacacattcctacagcaattgcataccttcgggcggtcgggactgggcaatgcagctacaacatcctcgcctaaagttatgcaat
tcgagcgacaaatgttgccgtgttagggcttttttgtgataatagtcgttttttttgtcctctcgcttatcaaactctatcaacggaggaaa
tccattttcgctacaatgcctacagctcaagtttcaaggtcaatcgagcgggtggggatcaactttttttattcattttgctaacgcccca
tcaacaaattctatgttctcaatggcaaagattactgcccgcaccaatcgcccaacgaaacggcaaaagaaaagcgacgattatga
35 agatgtccaaaccattgcccgcccgacgctttatctgatgatttgcgggatggcttttacttgtctgctactttcaggcacaaaaggaa
atgaaaccagcgcaggctcgtttgccggcttgcggaggttcttcaggcactgaggctgagtacttaaatcgaacgattttttacgattc
tggatccagttttatgatgtggcctgcattacagtggcaattataccctgatgttcatttcattgcattttgtaagtttgtgctggtaacg
cccgtaacgattaattcttttcaaagagattctttcaaagagattcaaaatgtgtataacaaatgctaacgaatggaccgtacttgg
agggttgcggaaagtaacgttttaaaatattcatcacaatcctctgcaaacttgtgcttaattaattggtgcacaataagtttaaact
40 gtggcggcagatgtgtcgctgtccgcttccttccttcccagcaagctcgtgcgaaataatttattccatcattttaatacagccgtttgtg
cattttaattagcaaagcaatataaaaagcagctaaccatccccattaaaacaaagtgcttccgggcccaattgttatggcggtgga
aagtaatggttttaccagtggaagtgtcctttcccatcgtgggtacttcgcgatattcttgtcttatacaagtgcatacagaaaaaaa
ggacaaatcctccttgctatggtctaaggccagcttcggtaccgcttccgcttcgggatgtcataaagtttgatgggtgtttttaacatt
acttccgctcttaaccacctaatggacttttcatgcttgagctaaagttaaaccagccaccagcggtacgcaccgagccacggttgatt
45 tcggcggcggcctcatccccagtttttgcgccaccaatattgccttcattaatctgtaccctcggagcgttagggcccgcggacgagtcct
cgttgtaatgcaccgccatgccacgggacgggataatccgttgggacggcgcgaaagcgactatcgcggacggattggttcgaccg

tgctacaacacattttatgcttcacagatttacttcctgctgttttcgatggtccagagcaacctcgcggatgtcatgttctgctcctggt

tgctgctagcctgcgagcagctgcaacacttgaaggtaggtacggtagcaaacgtggttgtctttacatccgcgtgcagcattatcct

tatcgacgtgtagtgttaacggtaaaagaggaagcgataaaaaagcaacattctctcacaccctcgatctctctttattttctctctct

ctctctctctctctctctctctctctctctctctctctctctccatctcctcgggcag*GGTATTATGCGATCGTTGATG*

5 *GAGCTTTCGGCCTCGCTGGACACCTACCGGCCCAACTCTTCGCAACTGTTCCGAGCAA*

*TTTCAGCCGGTTCCAAATCGGAGCTGATCATCAACGAAG*gtatgtgaaacgtgtgctcgtggcagacg

gactcaaagagagcataacacaatccctggtagttcatttcaatgaccttaacactcggcaagctaagcgagacagtggggacag

tgagaaagagagaacaagaaaaaaaaccatcatccgtacgacatcatcgctacgtaccggtatttcaggatgaggaaataaaac

gctaggggaatgaaagtgcgacagaatgataaaacaatccccacccaggcccccagcctggacgaacggatgtagtgtgcgaagc

10 gagcaaaaaaagtcaaataaattgaagtttaaaaatagattttccccgtccatccgtggtggagcgtaaagcccggcggacaactt

cgagcacggcgaccgtgcacagtactgtgccacagttgtagggacggataagctccgttcctttttttatcctttttttttggagatttgt

ttgcgttcgcatcgttagacgagcttagtgccgtgttgctctaattgctatttattataaagcgcttccaaatagaagatcggttctctc

catttaatctatcgcgcctgtacgcctgaaactatgcactgtgctgtgaaaccgtcaagctcgagcacgacgaatggcccaccgtacc

acgcccgtggtgcccaaagcgcaacgcgaattgcatgttaacaaaccttttgcctaccatccaatccgtgtgaaattgcccgctctcttt

15 ctctcttttgcgctttcggtgtatcgaacggttttgtcccttttttttactttgctcttgatctcttgctgtgctcactttcatctcatgttttgc

ctgacggtggtgggttttcgaaaaaagagcgatttcttctgcgtgtgtgtgtggtttttttttaaataaccgctccaggtcgtgttgaacg

ctgcaggaccgatcggagctagtttattatcagctttagtgtttatcccacccatgccccacatcacgtctgtggagagtgggggaag

cttaagtccaatgtaatttaccgtgtttctgtcgttcgtcaccttcttcgtcgatggagattggtgcggttggcacgataaaagcccact

gcacgttacggaccgagggaaaggtctttttgtaggcctagcaacggtcctcattcaccgcatgggggtgtagctcagatggtagag

20 cgctcgcttagcatgtgagaggtaccgggatcgatacccggcatctccaacccacacaaaacgttttttaagaagattttttagggaa

gatattaacgcgggtacactgtgctcctctaagttggaagagtagatgagatgatgacaagggagaaggaacatgtgtacgtgttt

gatagcaaacacacaaacaacaatatcatctctgataataatctgatgtgtgatgtgtgtgtattgttgttatgctgcctttgccatct

tgtccctctctctcctgttcaactcctaaaagaattgtttggagtcctctcagttcctcgtaaagatcctttcgagattcttcttttcttttt

attatttattccacgagcctctgacataagtagccttccgcttatttcttctccttgcacttgtcagttccgtgtagagcgtcattttgag

25 gtttacacatttcccaccgacgcctgattgttacattgtcatctacattgctttccgtttaccgttccgccctttttttttttaacgctaccaca

g*AAAAGGATCCGGACGTTAAGGACTTTGATCTGAGCGGCATCTACAGCTCGAAGGCGG*

ACTGGGGCGCCCAGTTCCGTGCGCCGTCGACGCTGCAAACGTTCGACGAGAATGGCAG

GAACGGAAATCCGAACGGGCTTACCCGGAAGCAGGAAATGATGGTGCGCAGCGCCATC

*AAGTACTGGGTCGAGCGGCACAAGCACGTTGTACG*gtaggtatggtaatttctaaggtgtggtgtaaag

30 cctccaggttccatgaaaaagggatactttaccacagtaagagtttgtttgctggacttacattctttggagcattgtttggtgttgtg

ctgaaaccggttgcaatatcgttttgcgaagaaattatgtgtaaagcgtattacaatctcattcctctgttaatctgtaccaattgtgtc

agccccgaccgaaagcaggcctaattcgtaccagaaaaaccacaagctgtttgtaagcatcgatacgcccgaagctttcaatccagc

caaggcgccacctactattgacgtgacttttttgcacgttcacactctccctctcccattctttctataaccaatcgtcgctcagccagcat

cgcccggagtgaagtttttatttgaacgatatcacccgtatcgattttccactaaacatgcttaaatcgtttcacaaagctcccccaaa

35 atcccatttcaccaatccaccaatttgaagtccgtcgtcctttgtgtccttgtgtttgtgtgtttgtgtgagctggagacatggggggagt

gagtaaccgaacaacctcttgccgctgcttcacgatatcgaacagcaccaagataagcatcccttttttccctagccgatgtctccgata

tctcgattccgcttccagcgaggcaaagaaaaaggcgaactggctgacctcacccggggcgaggaaaaagcgtagggattacgtc

gagcagcacgagttgtgattcttcttcttctggttccataaatcgctgacggtttccattaccgcctgcggagtgcacacacgtgaag

ggaaagcgaaaacgtttagattccagcagcaacggcagcaccagaagcagcagcagcgcggcaaattgaatcatcctgacgcgat

40 gagttgtctgggttttcgggtcggtggcttacagcaccacaccatctgctgcagctaatacagctgtaaatttcgttagacatagactt

gattttacaatattacacacacacttacacacacagctatagatttgtcgcttggcgtatggctctgtacggcgtgccgtacatgccgc

gagccgtgttgctgctggttgcgatacggatcacgtccgattcgattcagcctgcgtgttttttggtgaagatccttatcggtgacccact

ttcagtgtgtcgagagcgagggtcactatggcgcctgtcagttggaaagctaggctcgattcaaagggccattgtgccagtgttcttt

ttaagatagcgataagctttgatcgaaatagtaaatcaaacattgtttctttttcctattccaaactgttgccaacctcattattacg

45 tttttgcagcgggtgtatagtaaattgcatactttaaggcgtgattttcaaatgtagcgttccgtatgcagaaacgccatggattatgc

aatttaaacaatgctgcttccttaacattcaaataacggcttattaaggaactttttgtgcaatttgtttttaacagcaaatagttagc

tcagaacgatcacatttagtatcgcttcaacaaagaactcttttaaacacacaatttgtaatgccattccctcgagaaagtttcttgtc

agtcctcctctgcatcacagcaacaaccaaacctgctcatgtttcctgctcgtttcctagctgttttgaacgttatttccgattcctgtgct

tgcccgcttttcttacaatcaaccacaatggttcagatttcgctcttatttttattgacccactgctttcgtgctgaagcccgtggaaacaa

tgcgccaagctcagcatccagccatgcatgtaaaatgagccacgcgacagattttagacatcgctttcgctctgcaccggaggtggtt

ttattcttgtttccgattcccacgtccattcgtcctgggtccgtccgccgggcccgaaaccgtaagccgtgcggggaattacgcaatcga

aacgagccagaaaatgagcacgccaaatgcaaagaaaatcccctttttgagtggtgctcctgccaccactcatctccccaactggtgg

5   gtgaaaaaccttgtgcgcccccttctctttccagaaaaaaaaacgcctcgctcgcacaaaaacatgctcgcccggtgaagctgcgtatgt

cgcagaagctcaaaccaacgccgccagcaagcatcaacaatttctattcaaacacccaacgcagcgcccaaaccggggtgcactgta

ctcagtagcgaagatgctcagattgtcccgtgcgctgctttcgatgcccgtttcggagcgggaagccatcgcttgccaacgttggcgat

gtcttttagccgtggatttgaattttctgaatatcacaggcgggcgcggtttgcctgcaaggttgttgcttcccacacgagcattgcttt

ccgtaccgcggtggggcgagttttcaacgcaaccttctacaagcaacgccacaacgcctgggagcgatatttaacagaaacaagaa

10   catcccgaacttcagcacatgccgtgatttgcctgttggaaaagcttttgtgagcgtgtgagttgaacgagctctattttcccagcgat

gggtggcatttgtgtggcatgctatcgtcagctttcttgaatctttacctctccattcgcctccattagtacacgcgtatggaaaatgg

gtgcaacggatcagaacggattttccgcgacagacttaataaagggaaagcaacgcgtttttttgcatgtgtagtgtttatgagcttt

atgccgttactttgcaattaaaaatagcaaaaaataacagtttttttttgtaagcggattacaaagaatgtatcagaatattacgtg

aaacattcatttcatgctgttaacgctcaaatagaatagtttttgtaacacggattgcataccttgccggtatcggttacattttcgccta

15   acagtatgcaatctgtttagctttgttgtttaatgactgcgttggtagtacaatatttatttacaccgcgtaatttatctcacaaattgc

aaaaaaatgtcaatctgtatcgattattcacacaaatcagatcccggaaccagtgtagcccaatgtgctcttattgaattaccacga

acaaatcaacctgatgcccgggtccgttggcaaacagcttgcgccgaagccgctcagtgtttcgtgcactaccgtgctgccattttgct

gccctcatcgaacagataaacagaagggcaactcttgtgagcatcgcaatgcccgtctgaagttccgtcgaaaatgggcctaaattc

aatttgacgcatttacccgcgaacaattgcgcgaaggctgtcaagtgtgttccacgaactgcgacaacaagcacacacacaaacac

20   aaatgttatcgtttcggcatgtttctcggtacaaagcgtgtggcgctatgtggcatgccgattcccagacagagtgatcgatagtaaa

tgtagcctatccggtagcattcaatttccttttctatcctcgcaaacaaagcccattctggggaggcgtggtgaagctttcaaaggcat

tgtgaaacaaatgtcctggttcggagggatgctggggaaagcaaacacggtgccgccatcgctgctaccgtcaatcgatcatgcatg

atgtgattaatatttgtgttattcacctgcgtatctatgcgtccgtcgtgtcgttcggatttccggaagtcaaggaaaaagcgactcca

tttgggattggttttttgcagcgaaaaatcaaaacattcgcacaaaaccgtcctccatttcaaatgcctacacttgtcactgtatatctct

25   ctttctctcgttttgccacgttgcag*TCTCGTTTCAGCAATCGGAGATACGTACGGTCCTGCCCTGCT*
GCTACACATGCTGACCTCCACCATCAAGCTGACGCTGCTCGCCTACCAGGCAACGAAA
ATCGACGGTGTCAACGTGTACGGATTGACCGTAATCGGATATTTGTGCTACGCGTTGG
*CTCAGGTTTTCCTGTTTTGCATCTTTGGCAATCGGCTCATCGAGGAG*gtacgtgcgctcggcgtg

ttgccgtgggaaagcattctccctgccccatatcgcttcattctcccagatcacacatttgcatcacaaagccagcacacttttgcttcg

30   ccgctgccatctcggcttctgaatgttttcacttctcccatacttctcccgtgcag*AGCTCATCCGTGATGAAGGCGGC*
CTATTCCTGCCACTGGTACGACGGGTCCGAGGAGGCAAAAACCTTCGTCCAGATCGTT
TGTCAGCAGTGCCAGAAGGCGATGACTATTTCCGGAGCCAAGTTTTTCACCGTTTCGC
*TCGATCTGTTTGCTTCG*gtaagtgtagcctggtggctggcacagaacaggctggcaaaacagggactttggctctagc

ctgatgggtggtatatgtgtgtctattttttgctaccattctcgcatcccttcctttccag*GTTCTTGGAGCCGTTGTCAC*

35   *CTACTTCATGGTGCTGGTGCAGCTGAAGTAA*acagccgtggcccggaaggatgtgtttttttcgctcgttcg

gttgtttgtttgtgcacactttctcttggacattttctctactgcaaaggtttaacaaacagcaacaacaaataatcccaagttttctttt

acagatctttgcaaaatgattagattttaatagattaacagtgcttgattatctgtcctgtagcaaccggggctgaagaacgttgatt

tggtaaaagtacaaaagggacgttggaaattgaaccaccagaagagtgatatttatgcaaagctcaccaagggaaatctatgtat

gtgtgatttgcgctcatcaagcactgtatgtgcctttcaactagtgcagcaataaagagtacaaatgtttcttagcgcaccgtacattg

40   tcgtttcggcgttttaaccgttgttgataatacacaaaagatgataaaaataaataataacaaaatgttaatatgagtaagtacta

aatagagaaatcgttttagtatgatcataccctccaatcatttgtttgaaattaactttaattttaactcaaattaaaccgatgtttttact

ttctgtgagaattattgtggaagaacttaatggaagtataattaaattgattgctaactttatgcgtttttcaatttacgaacgctagt

cttcaaacatcgcttcaaaagtattactaccacattattcatttacttatagttatatttattgcctcttcatctttccatggccagaact

actgcagaaaagcttctttttttgctcgctttccgatggttggttggacgaagttggtaacaaacggcaagcaattagcataaactatt

45   ttcgcatcgagatggaaatgaatgtaccactagaaccgagtgaaatgaattacttttcaacttgcacgccaaaaccattatctaaag

tacgcacaacttaaaaacaaaccccaaattgtcgtccacccttcattccactttcttgctacactttccgaccgagttctgtagcgccag

cagcaaaaaaatacatataaaaaccttcatcactcaagctgtatcgagccagcgtgggttgtgtttgactgtgctgtgaaagaaaga

agaaaaaaaaaacacttccacgggaagctagcaattggaaatgcataaattaaccggaagaaattcgcaaaaccccgcaccgac

gtaccgcaccgcatccgtaccgataccggaacaaacggtgtgcgcgaaagaatccgctagcagccccactggcacgggtatttgctt
ttggttctgtgtttttcttccactggtttgggtgcctgggcgaaggctagctcggctactttcccggggccgcaattttctgcagcccaag
gcggcgtgctcgtggggccaaaagaat

5

Figure 9 continued